

<https://helda.helsinki.fi>

Valintaverkkoanalyysi konekäännöksen jälkieditoinnin tarkastelun apuna

Salmi, Leena

2020-05

Salmi , L & Koponen , M 2020 , ' Valintaverkkoanalyysi konekäännöksen jälkieditoinnin tarkastelun apuna ' , MikaEL Kääntämisen ja tulkkauksen tutkimuksen symposiumin verkkojulkaisu , Vuosikerta. 13 , Sivut 58-74 .

<http://hdl.handle.net/10138/320216>

unspecified
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Valintaverkkoanalyysi konekäännöksen jälkieditoinnin tarkastelun apuna

Leena Salmi, Turun yliopisto
Maarit Koponen, Turun yliopisto & Helsingin yliopisto

Abstract

This paper presents an application of Choice Network Analysis (CNA; proposed by Stuart Campbell in 2000) to analyse Machine Translation (MT) post-editing (PE) data. The data were collected from 33 translation students in a PE experiment where the students post-edited a text machine translated from English to Finnish. The text was a combination of neural (NMT), rule-based (RBMT) and statistical machine translation (SMT) output.

CNA has been presented as a method for collecting data on the mental process of translating (or post-editing), to be used either instead of or in addition to experimental process studies. The method consists of collecting translations of the same text from different translators and building a network of the different options they have used to translate a particular item. The basic assumption is that when an item is translated in the same or a similar way, it requires less cognitive effort than an item that produced several different translations.

In this paper, we use CNA to analyse a subset of the PE data, items that are repeated in the post-edited text, to compare differences both between editors and between MT engines. We also discuss two different approaches to presenting the choice networks.

Keywords: Machine Translation, post-editing, Choice Network Analysis

Avainsanat: konekäännös, jälkieditointi, valintaverkkoanalyysi

1 Johdanto

Konekäännöksen käyttö käännösosalalla on yleistynyt viime aikoina myös Suomessa konekäänninteknologian kehittyessä. Ammatillisissa yhteyksissä konekäännöstä voidaan hyödyntää osana kääntämisen työnkulkua siten, että konekääntimellä tuotetaan raakaversio, jota kääntäjä sitten korjaa ja muokkaa tarvittavalla tavalla. Prosessia englanniksi kuvaava nimitys *post-editing of machine translation* (PEMT) viittaa konekääntämisen jälkeen tapahtuvaan toimintaan ja esiintyy usein myös ilman konekäännöksen viittaavaa osaa (*post-editing*, PE). Suomeksi käytössä on vastaavasti näkynyt nimitykset *post-editointi* tai *jälkieditointi*. Käytämme tässä artikkelissa muotoa (konekäännöksen) **jälkieditointi**. Englannin kielen *editing* tarkoittaa tekstien muokkaamista yleensä, mutta ihmiskäännöksen korjaamisesta on käytetty englanniksi nimitystä *revi-*

sion, jonka suomenkielinen termi ei ole vielä vakiintunut. Konekäännöksen käyttö Euroopan käännösmarkkinoilla yleistyy, mutta ei vaikuta vielä kovin vakiintuneelta. Käännösalan järjestöjen käännösalaan tarkastelemaan kyselyyn vastanneista yrityksistä vain noin 20 % ilmoitti konekääntimien olevan säännöllisessä käytössä, itsenäisistä kääntäjistä vielä harvempi (ELIA ym. 2019). Laatuongelmien vuoksi konekäännöksen käyttö ja jälkieditointi onkin ollut Suomen käännösmarkkinoilla vielä melko vähäistä: vuonna 2017 tehdyssä selvityksessä Suomen 12 suurimmasta käännösalan toimijasta kolme ilmoitti tarjoavansa konekäännöksen jälkieditointipalveluja asiakkaille ja kaksi ilmoitti kehittävänsä kääntimiä (Dranch 2018). Jälkieditointipalvelujen laajuudesta tai kysynnästä ei ole selvää kuvaa, ja kääntimien kehitystyö on painottunut pääosin englanti-suomi- ja suomi-englanti-kääntimiin. Konekäännin ja jälkieditointi on otettu osaksi käännösprosessia esimerkiksi valtioneuvoston kanslian kielipalveluissa syksyllä 2019 EU-puheenjohtajakauden yhteydessä (Kuhmonen 2019).

Jälkieditointi on selvästi yleistymässä, ja siksi se sisältyy myös tulevien kääntäjien koulutukseen. Tässä artikkelissa tarkastelemme aineistoa, joka käsittää kääntäjäopiskelijoiden editoinnit alun perin englannista suomeen konekäännetyistä tekstistä. Analyysissä sovellamme jälkieditointiin ns. **valintaverkkoanalyysiä** (*Choice Network Analysis, CNA*), joka on Stuart Campbellin (2000a ja 2000b) käännösprosessin tutkimiseen ehdottama menetelmä. Kääntäessä tapahtuvaa kognitiivista prosessia on tutkittu empiirisiin menetelmin ja tarkasteltu paitsi sitä, mitä itse prosessissa tapahtuu, myös sitä, mitkä seikat tekstissä aiheuttavat kognitiivista kuormitusta. Valintaverkkoanalyysissä tarkastellaan prosessin sijasta lopputulosta. Siinä verrataan toisiinsa eri kääntäjien samasta ilmauksesta tekemiä käännöksiä ja pyritään saamaan selville, mitkä lähdetekstin ilmaukset tuottavat useita erilaisia vaihtoehtoja ja ovat siten kognitiivisesti kuormittavia. Konekäännetyjä tekstejä ja niiden jälkieditointia tutkittaessa on havaittu, että jälkieditoijien valinnat eroavat toisistaan. Tässä tarkasteltava konekäännetty aineisto on tuotettu kolmella eri konekääntimellä. Tarkastelemme samasta konekäännetyistä ilmauksesta tuotettujen jälkieditointien välisiä eroja ja lisäksi sitä, miten erilainen konekääntimen tuotos vaikuttaa tekstin muokkaamiseen. Lisäksi esitämme kaksi erilaista tapaa tehdä valintaverkkoanalyysi.

2 Valintaverkkoanalyysi tutkimusmenetelmänä

Stuart Campbell (2000a ja 2000b) esitteli valintaverkkoanalyysin mahdollisena menetelmänä käännösprosessin tutkimiseen. Siinä verrataan toisiinsa eri kääntäjien samasta ilmauksesta tekemiä käännöksiä (kaikki kyseiselle ilmaukselle mahdolliset käännökset). Oletuksena on, että jos ilmaus on käännetty samalla tavalla, kyseessä on helposti käännettävä ilmaus, joka aiheuttaa vain vähän kognitiivista prosessointia. Jos taas vaihtoehtoja on paljon, kyseessä on vaikea tai hankala kohta, joka aiheuttaa paljon kognitiivista kuormitusta.

Campbellin (2000b: 32) mukaan eri kääntäjien laatimien kohdetekstien versioita vertaamalla voidaan saada tietoa käännösprosessin aikana tehdyistä valinnoista ja muodostaa yksittäisten tekstien ja kieliparien perusteella hypoteeseja testattaviksi muissa

teksteissä ja kielipareissa. Eri kohdetekstiversioita vertaamalla havaittuja kohtia, joissa kääntäjien välillä syntyy eroja, voidaan ajatella päätöksentekopisteinä, jotka haarautuvat (mas. 35–36). Mitä enemmän eri ”haaroja” verkkoon muodostuu, sitä enemmän päätöksiä kääntäjä joutuu tekemään ja sitä vaikeampi teksti(kohta) on kyseessä (mas. 38). Hale ja Campbell (2002) katsovat, että vaihtoehtoja tarkastelemalla voidaan arvioida paitsi tekstikohtien vaikeutta myös käännöksen tarkkuutta. Toisaalta voidaan ajatella, että käännösvaihtoehtojen runsas lukumäärä ei välttämättä tarkoita ongelmallista kohtaa, vaan saattaa kuvastaa monitulkintaisuutta, joka voi antaa kääntäjille mahdollisuuden suurempaan luovuuteen (Ketola 2018: 70–71).

Valintaverkkoanalyysiä on sovellettu esimerkiksi metaforien käännösratkaisujen luokitteluun (Campbell 2000b), tiettyjen vaikeiden rakenteiden käännösvaihtoehtojen tarkkuuden analysointiin ja arviointiin (Campbell 2000a; Hale & Campbell 2002), käännösongelmien tarkasteluun kääntäjäopiskelijoiden yhteistoiminnallisissa käännöstehtävissä (*Collaborative Translation Protocol*) (Pavlović 2009), kuvia sisältävien tekstien tekstien käännösongelmien ja -ratkaisujen analysointiin (Ketola 2016, 2018) sekä käännösten kirjaimellisuuden määrittämiseen (Schaeffer & Carl 2014). Tutkittavana kohderyhmänä ovat pääosin olleet kääntäjäopiskelijat, joskin Campbell (2000a) ehdottaa, että valintaverkkoanalyysillä vaikeiksi tunnistettujen kohtien analysointia voitaisiin soveltaa esimerkiksi auktorisointia varten tehtävässä arvioinnissa erityisesti käännettäessä B-työkielelle. Ihmiskääntäjille mahdollisesti syntyvien käännösongelmien lisäksi menetelmää on käytetty myös konekääntimille vaikeiksi tunnettujen lähtöteksti-ilmiöiden analysointiin (O’Brien 2005, 2006).

Menetelmän lähtökohtana on, että monen eri kääntäjän kohdetekstivaihtoehtoista muodostuva verkko kuvastaa teoreettisesti, millainen kognitiivinen malli kunkin kääntäjän käytettävissä on hänen kääntäessään (Campbell 2000a: 215). Kuten Nataša Pavlović (2009: 86) toteaa, kohdetekstistä muodostuvasta verkosta ei kuitenkaan pysty jäljittelemään, mitä vaihtoehtoja kukin kääntäjä tosiasiallisesti pohti, mistä nämä vaihtoehdot olivat peräisin tai miksi hän päätyi lopulliseen versioon. Barbara Dragstedin (2012: 83) mukaan on mahdollista, että yksittäinen kääntäjä on kuitenkin harkinnut vain yhtä vaihtoehtoa ja käyttänyt kohdetekstissä tätä ensimmäistä vaihtoehtoaan, jolloin ratkaisu on ollut tälle kääntäjälle helppo. Toisaalta Anne Ketola (2016: 80) nostaa esiin myös mahdollisuuden, että kääntäjä on harkinnut ratkaisuja, jotka eivät näy lopullisissa kohdetekstiversioissa. Saman toteaa Sharon O’Brien (2006: 177) jälkieditoinnin osalta. Tietyn ryhmän käännöksiä tarkastelemalla havaitut vaihtoehdot eivät myöskään välttämättä kerro sitä, mitä vaihtoehtoja joku muu ryhmä tai joku ryhmän ulkopuolinen olisi lähtötekstin perusteella muodostanut (Pavlović 2009: 65). Campbellin (2000b: 32) mukaan käytettävissä olevien vaihtoehtojen joukkoa voidaan tarkastella kokonaisuutena vasta, kun vertailtavien käännösten määrä on riittävän suuri. Toisaalta suuren kääntäjäjoukon käännösten analysointi on käytännön tasolla työlästä, ja siksi Ketola (2018: 66) ehdottaakin valintaverkkoanalyysin soveltuvan parhaiten melko pienten ryhmien tarkasteluun. Edellä käsitellyistä tutkimuksista useimmissa analysoitujen rinnakkaisversioiden lukumäärä on ollut 8–11.

O’Brien (2006: 178) mainitsee valintaverkkoanalyysin puutteeksi myös sen, että siitä ei näy, miten kääntäjän tai jälkieditoijan aiemmat valinnat vaikuttavat tarkastelta-

van tekstikohdan vaihtoehtoihin. Tällaisia valintoja voivat olla esimerkiksi termit tai lukijan puhuttelu. Yhden segmentin työläyttä tarkasteltaessa olisikin syytä ottaa huomioon myös kääntäjän tai editoijan aiemmat valinnat (mts. 178).

Valintaverkkoanalyysillä esiin nostettuja mahdollisia käännösongelmia on pyritty yleensä myös yhdistämään muunlaiseen käännösprosessista kerättävään aineistoon (tällaista saman ilmiön tarkastelua eri aineistoissa kutsutaan **triangulaatioksi**). Tarkempaa tietoa mahdollisiin käännösongelmiin liittyneistä pohdinnoista ja ratkaisun valinnasta on pyritty saamaan esimerkiksi kääntäjäopiskelijoiden ryhmäkeskustelujen (Pavlović 2009) tai käännösraporttien ja ryhmähaastattelun (Ketola 2016) perusteella. Ketolan (mts. 92) tutkimuksessa kääntäjäopiskelijoiden pohdinnoista nousseet havainnot viittasivat siihen, että valintaverkkoanalyysissä tunnistettu runsas vaihtoehtojen määrä osoitti, etteivät he olleet ymmärtäneet kunnolla kyseistä kohtaa, joissakin tapauksissa ristiriitaisen kuvallisen informaation vuoksi. Valintaverkkoanalyysimenetelmässä oletetaan, että kohdetekstissä havaittujen vaihtoehtojen lukumäärällä on yhteys käännösprosessin vaikeuteen, ja tätä on myös tarkasteltu yhdistämällä kohdetekstin valintaverkkoanalyysiä kääntämisen aikana kerättyyn prosessiaineistoon, kuten näppäilyntallennus- ja silmänliikeaineistoon (ks. O'Brien 2005, 2006; Dragsted 2012; Schaeffer & Carl 2014).

Dragsted (2012) pyrki selvittämään, voidaanko valintaverkkoanalyysillä tunnistetun kohdetekstissä esiintyvän vaihtelun perusteella ennustaa prosessiaineistossa näkyviä kognitiivista kuormitusta kuvastavia piirteitä. Niitä edustivat hänen tutkimuksessaan silmänliikeaineistosta havaittujen fiksaatioiden (katseen kiinnittymisten) lukumäärä ja kokonaisfiksaatioaika, näppäilyntallennuksesta tunnistettujen taukojen pituus sekä kohdetekstin kohtien uudelleenmuokkaaminen. Valintaverkkoanalyysin perusteella Dragsted (mas. 87–89) tunnisti tiettyjä sanoja, joissa kohdetekstissä esiintyi joko paljon vaihtelua (kahdeksalta koehenkilöltä 5–8 vaihtoehtoa) tai vähän vaihtelua (1–2 vaihtoehtoa), ja vertasi näihin sanoihin liittyviä prosessimittareita. Sanoihin, joille esiintyi useita käännösvaihtoehtoja, kohdistui silmänliikeaineistossa useampia ja pidempiä fiksaatioita ja niiden kääntämistä edelsi näppäilyntallennuksessa pidempiä taukoja, mikä Dragstedin (mas. 95) mukaan tukee oletusta, että kohdetekstissä havaittu vaihtelu kuvastaa kääntämisen vaikeutta. Prosessidatan perusteella ei kuitenkaan ollut havaittavissa, että useita vaihtoehtoja saaneita sanoja muokattaisiin useammin, eikä tutkimuksessa siis saatu selkeää näyttöä siitä, harkitsiko yksittäinen kääntäjä useampaa kuin yhtä vaihtoehtoa (mp.).

Moritz Schaeffer ja Michael Carl (2014) sovelsivat valintaverkkoanalyysin perusajatusta käännösratkaisujen ”kirjaimellisuuden” määrittämiseen erilaisissa prosessitutkimuksissa kerätyn käännöskorpuksen (nimeltä TPR-DB) käännöksissä. He laskivat korpuksen sisältämien rinnakkaisten käännösversioiden perusteella kutakin lähtökielen sanaa vastaavien käännösvaihtoehtojen määrän, tarkastelivat tämän lukumäärän yhteyttä silmänliikeaineistossa ilmeneviin kognitiivisen kuormituksen mittareihin (kokonaisfiksaatioaika, fiksaatioiden määrä) ja vertasivat käännös- ja jälkieditointilanteita. Tulokset vahvistavat Dragstedin havaintoa, että vaihtoehtojen määrä korreloi fiksaatioiden lukumäärän kanssa ja lisäksi vaikutus oli jälkieditoinnissa voimakkaampi kuin kääntäessä (Schaeffer & Carl 2014: 35).

O'Brien (2005, 2006) puolestaan tutki, löytyvätkö samat hankalat rakenteet ja kognitiivisesti kuormittavat kohdat sekä valintaverkkoanalyysillä että prosessintallennusohjelmalla. O'Brienin mukaan joillakin hankalaksi arvioituilla rakenteilla, kuten gerundilla tai pitkillä substantiivirakenteilla, oli suurempi vaikutus jälkieditointiin kuin lyhenteillä ja demonstratiivipronomineilla. (O'Brien 2006: 172–173.)

3 Valintaverkkoanalyysin periaatteet

Valintaverkkoanalyysi on menetelmänä avoin. Campbell (2000b: 39) määrittää analyysissä syntyvän verkon muodostamiselle kolme periaatetta:

1. Verkon pitää kattaa kaikki havainnot, yhtään kohdeteksteissä havaittua vaihtoehtoa ei saa jättää pois vaikka kyseessä olisi analysoijan mielestä ”outo” ratkaisu tms.
2. Verkon pitää olla lingvistisesti perusteltu, esimerkiksi tarkasteltavia yksiköitä (”solmuja”) ei saa muodostaa niin, että ne ovat kieliopillisten yksiköiden vastaisia.
3. Verkon solmujen ja haarojen lukumäärä pitää minimoida.

Campbellin (2000b: 39–40) mukaan taustaoletus valintaverkkojen muodostamisessa on, että kääntäminen tapahtuu pääosin lineaarisesti alusta loppuun, vaikkakaan yksittäisissä kohdissa päätökset eivät välttämättä tapahdu siinä järjestyksessä kuin ne verkossa esitetään. Pavlović (2009: 86) huomauttaakin, että päätösten esittäminen Campbellin mallin mukaisena verkkona saa prosessin vaikuttamaan todellista tilannetta lineaarisemmalta prosessikaaviolta: analysoimiensa kääntäjäopiskelijoiden keskustelujen perusteella Pavlović totesi, ettei eteneminen ei ole läheskään aina lineaarista tai järjestelmällistä.

Analyysimenetelmä ei määritä, miten tarkasteltavat yksiköt muodostetaan, vaan kuten Hale ja Campbell (2002: 18) toteavat, kohdetekstistä ilmenevät vaihtoehdot voivat näkyä monella eri tasolla yksittäisistä leksikaalisista valinnoista laajempiin syntaktisiin rakenteisiin, kuten passiivin kääntämiseen. Siksi valintaverkkoanalyysissä tarkasteltavat yksiköt voivat olla erikokoisia yksittäisistä sanoista kokonaisiin lausekkeisiin (ks. Ketola 2018: 67). Jako riippuu myös tarkasteltavien kielten rakenteista. Kuten Ketola (2018: 67–68) kuvaa, kokonainen englannin prepositiolauseke saatetaan ilmaista suomeksi yhdellä sanalla, jolloin tarkasteltava yksikkö täytyy määrittää niin, että se on molempien kielten kannalta perusteltu. Menetelmä ei myöskään määrää, mitä kategorioita vaihtoehtojen analyysiin tai nimeämiseen käytetään, vaan eri analyyseissä voi esiintyä erilaisia luokitteluja tutkimuksen tavoitteiden mukaisesti. Esimerkkejä aiemmissa tutkimuksissa käytetyistä luokitteluista ovat metaforien käännösratkaisujen luokittelu pelkistävään, säilyttävään tai muokkaavaan (Campbell 2000b: 34), relatiivilauseen vastineiden luokittelu relatiivilauseisiin, parafraaseihin tai muihin ratkaisuihin (Campbell 2000a: 222), organisaatioiden nimien luokittelu sen perusteella, onko nimi käännetty vai transkriboitu mahdollisesti eksplikoiden (Hale & Campbell 2002: 20) sekä erilaisten käännösratkaisujen yleinen luokittelu alkuperäisen merkityksen säilyttävään tai merkitystä muokkaaviin (Ketola 2016). Metodologisen avoimuuden vuoksi

Ketola (2018: 69) katsookin, että valintaverkon rakentaminen itsessään on analyysin keskiössä, koska se edellyttää analysoijan pohtivan ja päättävän, miten vaihtoehdot määritetään ja esitetään.

4 Aineisto

Aineistomme sisältää 33 maisterivaiheen kääntäjäopiskelijan jälkieditoinnit: sekä Translog II -ohjelmalla (ohjelmasta tarkemmin Carl 2012) kerättyä prosessidataa että jälkieditoinnin tuotokset. Aineisto kerättiin kahdessa osassa: lokakuussa 2016 Helsingin yliopistossa (osallistujat P01–P16) ja maaliskuussa 2017 Turun yliopistossa (osallistujat T01–T17). Jälkieditointi oli yhtenä tehtävänä tämän artikkelin kirjoittajien opettamalla, konekääntämistä ja jälkieditointia käsitelleellä kurssilla. Osallistujista 31 ilmoitti äidinkielenkseen suomen ja kaksi ilmoitti olevansa kaksikielisiä (suomi ja venäjä / suomi ja englantia). Kaikkien keski-ikä oli 26,9 vuotta (mediaani 24 vuotta). Koehenkilöistä oli naispuolisia 21, miespuolisia 12.

Tehtävässä käytetty teksti oli lyhyt, yleistajuinen BBC:n verkkosivuilla julkaistu teksti, joka oli käännetty englannista suomeksi kolmea eri konekäännintä käyttäen: sääntöpohjaista (RBMT), tilastollista (SMT) ja neuroverkkokäännintä (NMT). Lyhenneet tulevat käänninten englanninkielisistä nimityksistä (*Rule-Based*, *Statistical* ja *Neural Machine Translation*). Englanninkielinen lähdeteksti ja suomenkieliset SMT- ja NMT-käännökset olivat peräisin vapaasti saatavilla olevasta korpuksista, joka on koottu ACL First Conference on Machine Translation (WMT16) -konferenssin yhteydessä järjestettyä konekäännöksen laatuarviointikampanjaa¹ varten (kampanjaa ja aineistoa kuvailevat tarkemmin Bojar ym. 2016). Sääntöpohjainen käännös oli tuotettu suomalaisella Sunda-kääntimellä (ks. <https://www.sunda.fi/>).

Lähdeteksti koostuu 27 virkkeestä (385 sanaa) ja sisältää ohjeita, miten lähettää materiaalia tai uutisvinkkejä BBC:lle. Valitsimme tekstin siksi, että se oli yleiskieltä eikä sisältänyt erikoisalan termejä, joten se olisi suhteellisen helppo korjata, ja koska se sisälsi joitakin toistuvia ilmauksia, joiden jälkieditointia olisi kiinnostavaa tarkastella.

Koska halusimme, että kaikki osallistujat jälkieditoisivat kaikkien konekäänninten tuotoksia, jaoimme tekstin kolmeen osaan siten, että jokaisella osallistujalla oli kunkin konekääntimen tuottamasta tekstistä 9 virkettä, taulukon 1 mukaisesti:

Taulukko 1: Konekääntimien tuotoksien jakautuminen osallistujien jälkieditoimissa tekstiversioissa.

	Tekstiversio 1	Tekstiversio 2	Tekstiversio 3
Virkkeet 1–9	RBMT	SMT	NMT
Virkkeet 10–18	NMT	RBMT	SMT
Virkkeet 19–27	SMT	NMT	RBMT
Osallistujien määrä	13	9	11

¹ Korpus on ladattavissa verkkosivulta <http://www.statmt.org/wmt16/translation-task.html>

Osallistujille annettiin kirjalliset ohjeet, joissa esiteltiin lyhyesti Translog-ohjelman toimintaa ja kehoitettiin käyttämään tietolähteinä ainoastaan lähtötekstiä ja konekäännöstä. Tarkemmat editointiohjeet olivat standardiin ISO/DIS 18587:2016 sisältyvät ns. kevyen jälkieditoinnin ohjeet:

- Käytä raakaa konekäännöstä niin paljon kuin mahdollista.
- Pyri tuottamaan merkitykseltään oikea, suomen kielioppisääntöjen mukainen käännös.
- Tarkista ettei virkkeestä puutu informaatiota eikä siinä ole mitään ylimääräistä.
- Korjaa lauserakennetta, jos merkitys on väärä tai epäselvä.
- Noudata oikeinkirjoitus- ja välimerkkisääntöjä. (Käännös LS & MK.)

Aineistosta on aiemmin analysoitu 5 koehenkilön otoksen osalta sitä, millaisia muutoksia jälkieditoinnissa oli tehty (Koponen & Salmi 2017), ja koko aineiston osalta editointiprosessia sekä eroja RBMT-, SMT- ja NMT-kääntimen käännösten välillä (Koponen, Salmi & Nikulin 2019). Vertailussa havaittiin, että RBMT-kääntimen tuotosta editoitaessa tehtiin eniten poistoja muihin kääntimiin verrattuna, mikä johtuu lähinnä siitä, että RBMT oli kääntänyt kaikki persoonapronominit, joita ei aina tarvita suomenkielisen tekstin sujuvuuden kannalta. SMT-käännökseen tehtiin eniten lisäyksiä, ja NMT-käännökseen korvattiin sanoja ja korjattiin sanojen taivutusmuotoja. Tulokset kertovat siitä, että SMT-käännin jätti pois sanoja ja NMT-käännin tuotti merkitykseltään virheellisiä vastineita. Editointiprosessissa ei havaittu käänninten välillä selkeitä eroja. Lisäksi tarkasteltiin tehtyjen muutosten oikeellisuutta, eli sitä vastasiko lopputulos lähdetekstin sisältöä ja oliko se kielipiillisesti oikein, sekä niiden tarpeellisuutta, eli sitä oliko muutos välttämätön lähdetekstin sisällön välittämiseksi. Esimerkiksi persoonapronominien poistaminen luokiteltiin tarpeettomaksi poistoksi, jos virkkeen sisältö olisi välittynyt ja kieli ollut kielipiillisesti hyväksyttävää muutosta tekemättä. Tässä analyysissä kävi ilmi, että tarpeettomia muutoksia oli eniten RBMT-käännöksen jälkieditoinnissa. (Koponen, Salmi & Nikulin 2019.)

5 Valintaverkkoanalyysi

Tässä artikkelissa tarkastelemme valintaverkkoanalyysin avulla sellaisia aineistossa toistuvia tai kääntämättä jääneitä ilmauksia, joiden kohdalla konekääntimien tuotoksessa oli vaihtelua. Tämän artikkelin puitteissa ei kuitenkaan ollut mahdollista verrata valintaverkkoanalyysin tuloksia prosessidataan. Kuten mainittiin, RBMT-kääntimen tuotoksessa pronominit *you* ja *your* oli johdonmukaisesti käännetty monikkoon tai teitittelynä, ja aina kun lähdetekstissä esiintyi pronomini, myös konekäännöksessä oli pronomini (esimerkiksi virkkeessä 21: *You* can download → *Te* voitte ladata). NMT-kääntimen tuotoksessa esiintyi sekä yksikön että monikon 2. persoona, SMT-kääntimen tuotoksessa vain yksikön 2. persoona, mutta pronomini saattoi molemmista jäädä pois (esimerkiksi virkkeessä 21: ”Voit ladata” sekä SMT- että NMT-tuotoksessa). Kuten mainittu, kaikilla osallistujilla oli kaikkien kolmen kääntimen tuotosta.

Analysoidut ilmaukset ja virkkeet, joissa ilmaukset esiintyvät, olivat seuraavat:

- *contribution* (virkkeet 4, 11, 12 ja 21)
- *breaking news* (virke 4)
- *We will publish your name as you provide it* (virkkeet 18, 25 ja 27)
- *unless you ask us not to* (virkkeet 18, 25 ja 27)

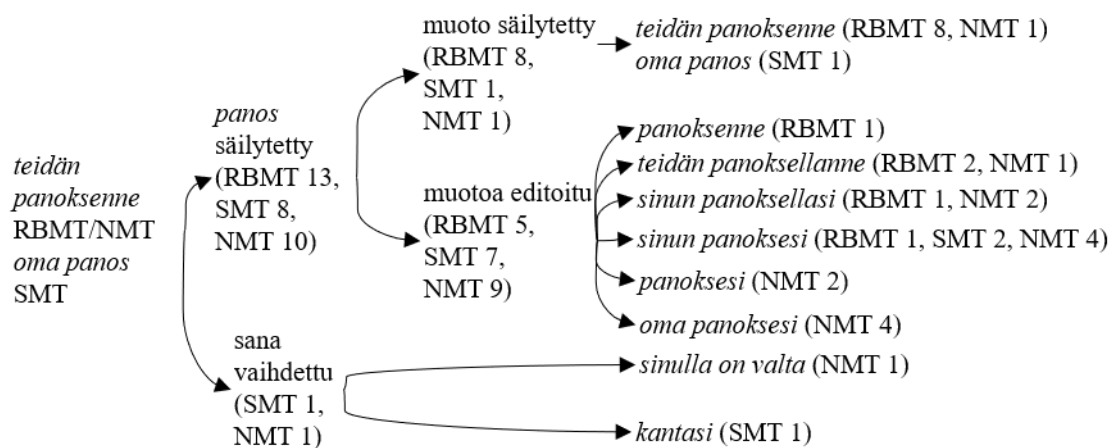
Lähdekieliset virkkeet ja niiden konekäännökset on esitetty kokonaisuudessaan liitteessä 1. Kaksi viimeksi mainittua ilmausta esiintyivät hieman eri yhdistelminä kolmessa virkkeessä. Näistä virkkeistä tarkastelimme lisäksi sinuttelun ja teitittelyn vaihtelua.

5.1 *Contribution*-sanana vastineet

Lähtetekstin *contribution* esiintyy neljässä eri virkkeessä (virkkeet 4, 11, 12 ja 21). Sanalla tarkoitetaan materiaalia, jota BBC:n lukija voi lähettää toimitukselle, esimerkiksi uutisvinkit, valokuvat tai videot. Termi on työläs jälkieditoida, sillä sille ei ole suomeksi yhdellä sanalla ilmaistavaa vastinetta. Konekäännöksissä esiintyi kaikkiaan kuusi eri ilmausta: *panos*, *avustus*, *maksu*, *jäsenmaksu*, *puheenvuoro* ja *osuus*. Näistä *panos* oli yleisin. Se oli RBMT-kääntimen vastine kaikissa muissa paitsi virkkeessä 11.

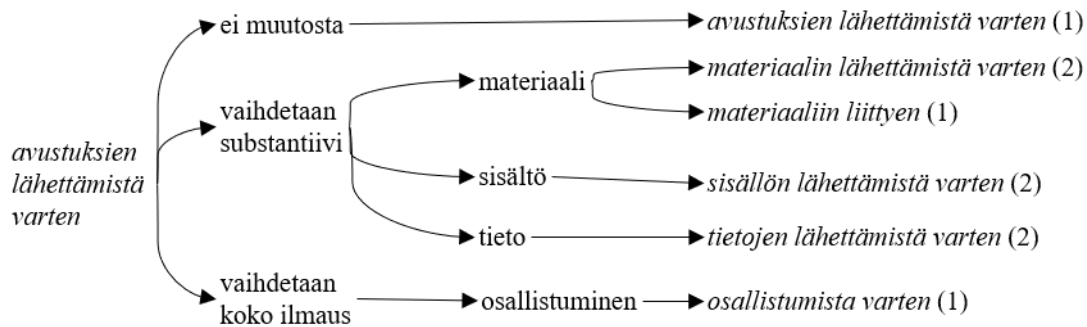
Virkkeessä 4 *contribution*-sanana vastine oli *panos* kaikkien konekääntimien tuotoksessa, ja sitä on editoitu melko vähän. Kuvassa 1 on sen valintaverkkoanalyysi Campbellin (2000b) ja Ketolan (2016) tapaan esitettynä.

Kuva 1: Valintaverkko *panos*-sanana editoinneista virkkeessä 4 (kaikki kääntimet)

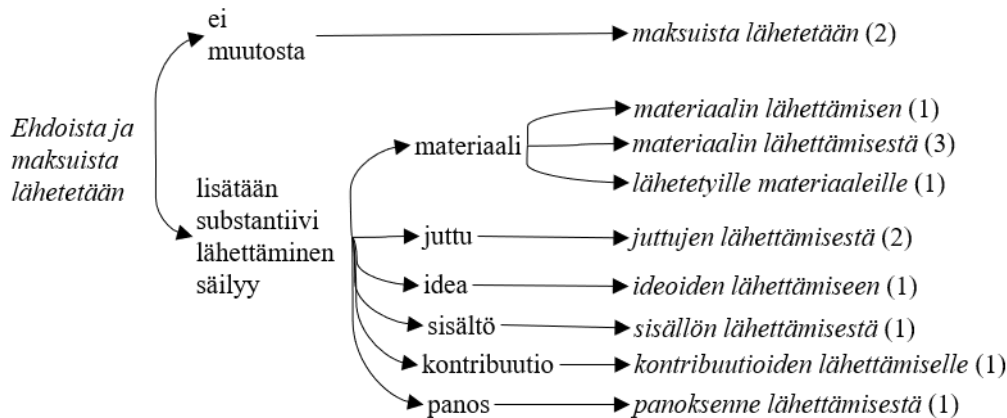


Virkkeessä 11 *contribution*-sanana vastineena on RBMT-kääntimellä *avustus*, SMT-kääntimellä vastine on merkitykseltään virheellinen *maksu*, ja NMT-käännöksestä vastine puuttuu. Tässä virkkeessä se on yleisimmin editoitu sanaksi *materiaali*. Kuvissa 2, 3 ja 4 on esitetty valintaverkkoanalyysit kaikista eri kääntimien *contribution*-sanaa vastaavien tuotosten editoinneista.

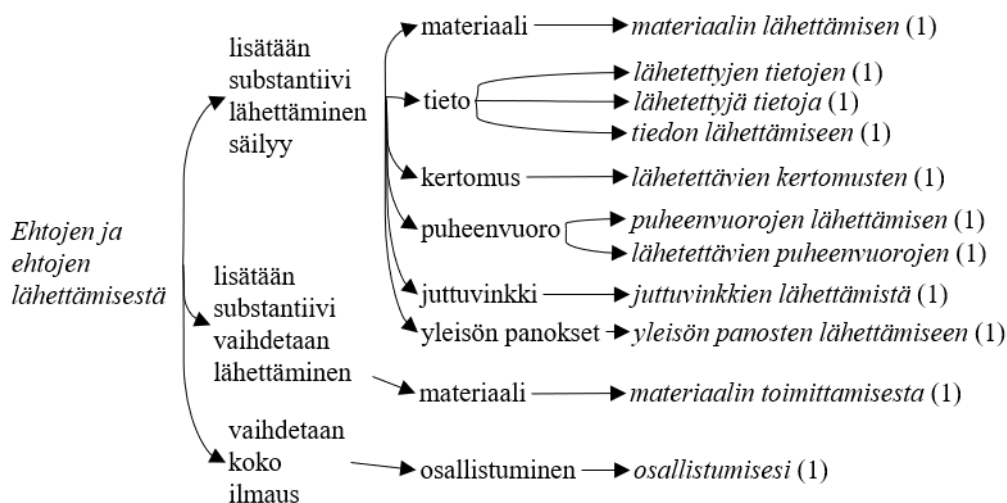
Kuva 2: Valintaverkkoanalyysi virkkeen 11 ilmauksen *avustuksien lähettämistä varten* editoinneista (RBMT-käännös)



Kuva 3: Valintaverkkoanalyysi virkkeen 11 ilmauksen *maksuista lähetetään* editoinneista (SMT-käännös)



Kuva 4: Valintaverkkoanalyysi puuttuvan ilmauksen editoinneista virkkeessä 11 (NMT-käännös)



Merkitykseltään virheellinen vastine SMT-käännöksessä (kuva 3) ja vastineen puuttuminen kokonaan NMT-käännöksestä (kuva 4) tuottivat enemmän erilaisia vaihtoehtoja (6 ja 8 kappaletta) kuin vastine *panos* (4 kappaletta), mikä viittaa suurempaan kognitiiviseen kuormitukseen. Siihen viittaa myös suora laina *kontribuutio*.

Virkkeessä 12 *contribution*-sanan vastineet konekääntimillä olivat *teidän panoksenne* (RBMT), *puheenvuoroista* (SMT) ja *jäsenmaksuissasi* (NMT). Esitämme seuraavassa valintaverkkoanalyysin tulokset tiivistetyssä muodossa:

RBMT: *lähettää teidän panoksianne* → editoinnit: *materiaali* (2), *tietojasi* (2), *sisältöjä* (1), *uutissisältöä* (1), *osallistua* (2)

SMT: *lähettää puheenvuoroista* → ei muutosta (1); editoinnit: *panos* (2), *tieto* (1) / *tiedot* (1), *puheenvuoroja* (1) / *puheenvuorosi* (2), *sisältö* (1), *ideat* (1), *vihjeet* (1), *työpanos* (1)

NMT: *lähettää jäsenmaksuissasi* → ei muutosta (1); editoinnit: *materiaali* (4), *jutut* (2), *panos* (2), *ideat* (1), *kontribuutio(t)* (1), *eri tileiltäsi* (1), *tuotelmus* (1)

Peräkkäisissä virkkeissä 11 ja 12 suurin osa osallistujista käytti samaa vastinetta: 20 käytti samaa sanaa molemmissa virkkeissä ja 13 vaihtoi sanaa (kolme osallistujaa jätti jommankumman tai molemmat virkkeet editoimatta, ja heidät on tässä laskettu sanaa vaihtaneiksi). Huolimatta siitä, että vastine *panos* oli käytössä kaikilla paitsi kahdella editoijalla virkkeessä 4 ja se oli RBMT-kääntimen vastine virkkeessä 12, yhteensä vain viisi editoijaa käytti sitä virkkeissä 11 (P08 ja P11) ja 12 (P08, P11, T03 ja T10). Lisäksi yhdellä editoijalla (T11) oli virkkeessä 12 ja 21 vastine *työpanos*.

Virkkeen 12 NMT-vastine *lähettää jäsenmaksuissasi* tuotti kolmella editoijalla kiinnostavan vastineen. Suora laina *lähettää kontribuutiosi* esiintyi samalla editoijalla (P05) kuin virkkeessä 11, mutta editoija P09 kehitti uudissan *tuotelmus* (lähettää tuotelmukseenne), ja kolmannella (P04) oli vastine *eri tileiltäsi*, jossa merkitys on muuttunut. Uudissan kehittämistä selittää se, että P09 jätti virkkeen 11 editoimatta, joten hänellä ei ollut siellä vastinetta jota toistaa, mutta P04 on edellisessä virkkeessä käyttänyt sanaa *materiaali*.

Virkkeessä 21 RBMT-kääntimen vastine oli *panokset*, SMT-kääntimeltä vastine puuttui ja NMT-kääntimen vastine oli *osuuksia*. Vastinetta *panos* käytti virkkeen editoimatta jättäneen P11:n lisäksi vain kaksi tätä käännöstä editoineista (P13 ja T04), jotka siis olivat eri henkilöitä kuin sanaa virkkeissä 11 ja 12 käyttäneet. Muita yleisiä vaihtoehtoja ovat *materiaali*, *sisältö* ja *tieto/tiedot*. Vastineen puuttuminen SMT-käännöksestä ei tuottanut suurempaa määrää vaihtoehtoja kuin muiden kääntimien tuotoksissa esiintyneet vastineet: SMT-käännöstä editoineilla 13 editoijalla oli erilaisia vastineita 4 (*materiaali*, *ideat*, *jutut* ja *sisältö*) ja kaksi editoijaa jätti vastineen kokonaan pois, kun taas muita käänninversioita editoineet olivat ehdottaneet kummankin kääntimen tuotokseen kuutta muuta vaihtoehtoa.

Vertaillaksemme erilaisia tapoja tehdä valintaverkkoanalyysi esitämme taulukossa 2 analyysin virkkeen 11 SMT-käännöksen editoinneista taulukon muodossa samoin kuin

O'Brien (2006: 165–167, Appendix H). Taulukkomuodossa esittämistä varten virke on ensin jaettava mielekkäisiin segmentteihin ja käytämme tässä samaa segmentointia kuin aiemmassa analyysissämme (Koponen ym. 2019: 72–73).

Taulukko 2: Virke 11, SMT-käännöksen editoinnit

ST11	Terms and conditions	for sending contributions	to the BBC		
SMT	Ehdoista ja	maksuista lähetetään	BBC.		
P11	Ehdot,	joita sovelletaan yleisön panosten lähettämiseen	BBC:lle		
P13		Puheenvuorojen	BBC:lle	lähettämisen	ehdot.
P15			BBC:lle	materiaalin lähettämisen	ehdot.
T02			BBC:lle	lähetyksen tietojen	ehdot ja maksut.
T03	Ehdot	materiaalin toimittamisesta	BBC:lle.		
T04		Osallistumisesi			ehdoista.
T05	Ehdot ja säännöt	tiedon lähettämiseen	BBC:lle.		
T09			BBC:lle	lähetyksen tietoja koskevat	käyttöehdot.
T11			BBC:lle	lähetyksen kertomusten	säännöt ja ehdot.
T15			BBC:lle	lähetyksen puheenvuorojen	ehdot ja säännöt.
T17	Ehdot ja maksut	koskien juttuvinkkien lähettämistä	BBC:lle.		

Taulukkomuoto mahdollistaa kaikkien osallistujien editointien vertaamisen allekkain ja yhdellä silmäyksellä. Vertailu kuitenkin vaikeutuu hieman silloin, kun osallistujat ovat käyttäneet eri sanajärjestystä ja erilaista syntaksia, kuten taulukossa 2: SMT-käännöstä editoineista 11 koehenkilöstä suurin osa (7) on muuttanut virkkeen sanajärjestystä ja syntaksia.

5.2 *Breaking news* -ilmauksen vastineet

Virkkeessä 4 lähdetekstin *breaking news* oli kääntynyt joko sanalla *uutinen* (RBMT), ilmauksella *rikkonut uutisia* (NMT) tai jäänyt englanninkieliseen muotoon *Breaking News* (SMT). RBMT-kääntimen tuotosta *uutinen* editoineista seitsemän jätti sanan samaan muotoon, kaksi editoi muotoa (*uutiset* tai *uutisesta*), kolme lisäsi alkuun jonkin määrään (*uusi uutinen*, *jymyuutinen* tai *suuret uutiset*) ja yksi yhdisti sen virkkeessä seuraavana olleeseen *featured item* -ilmaukseen (*minkälainen tarina tahansa*).

SMT-käännin oli jättänyt tekstiin englanninkielisen ilmauksen ja NMT-käännin kääntänyt sen virheellisesti. Kummankin kohdalla eri tavoin editoituja vaihtoehtoja oli yhtä monta kuin editoijia, mikä viittaa kognitiiviseen kuormitukseen. Suurimmalla osalla (29/33) editoijista oli kuitenkin tekstissään *uutinen* jossain muodossa; uutisen eteen lisättyjä adjektiiveja olivat esimerkiksi *tärkeä*, *tuore*, *merkittävä* tai *ajankohtainen* ja yhdyssanoja *jymyuutinen* tai *pääuutinen*. Kaksi editoijaa muutti merkityssisältöä (*päivän aiheista* tai *ajankohtaisasia*) ja neljä yhdisti sen *featured item* -ilmaukseen samoin kuin RBMT-käännöksen editoija edellä (esimerkiksi *minkälainen uutinen tahansa*). *Featured item* oli työläs editoida, sillä tehtävänannossa kiellettiin muiden lähteiden käyttö, joten jos merkitys ei ollut editoijalle tuttu, hän ei voinut tarkistaa sitä. Mikään käännin ei myöskään tuottanut sille hyväksyttävää vastinetta, vaan vastineet olivat *esitely kohta* (RBMT), *välilyksymys* (SMT) ja *roolihaamo* (NMT). Tämä selittää sitä, että yhteensä viisi editoijaa yhdisti ilmaukset.

5.3 Toistuvat ilmaukset virkkeissä 18, 25 ja 27

Virkkeissä 18, 25 ja 27 esiintyi kaikissa teksti ”We will publish your name as you provide it (unless you ask us not to)”, ja virkkeissä 18 ja 25 se jatkui: ”but we will never publish your”. Lause loppui virkkeessä 18 ”mobile phone number” ja virkkeessä 27 ”email address”. Tarkastelimme näistä virkkeistä paitsi editointien vaihtoehtoja, myös sitä, miten editoijat toimivat silloin, kun tekstissä oli toistoa. Kun lähdekielinen ilmaus esiintyy (lähes) samana kolme kertaa samassa tekstissä, olisi oletettavasti helpompaa editoida se vain kerran ja kopioida editoitu teksti muihin esiintymiin. Toisaalta osallistujille oli annettu ohjeeksi käyttää ”raakaa konekäännöstä niin paljon kuin mahdollista”.

Valintaverkkoanalyysissä havaittiin, että konekääntimen merkitykseltään virheellisesti kääntämät ilmaukset tuottivat eniten erilaisia vaihtoehtoja jälkieditoinnissa. Esimerkiksi virkkeissä 18 ja 25 lähdetekstin ilmauksesta ”as you provide it”, joka oli kääntynyt SMT-kääntimellä molemmissa virkkeissä muotoon ”kun se”, oli jälkieditoineissa yhtä monta erilaista muotoa kuin koehenkilöitä. Sen sijaan SMT-kääntimen käännöstä ”Julkaisemme nimesi” virkkeessä 18 ei muuttanut kukaan kyseistä käännöstä editoineista.

Koska vertailtavat tekstiosiot ovat tässä yhtä sanaa pitempiä, jaoimme ne neljään segmenttiin vertailun helpottamiseksi:

- *We will publish*
- *your name*
- *as you provide it*
- *(unless you ask us not to)*

Esitämme monisanaisten segmenttien analyysin taulukon muodossa samoin kuin O’Brien (2006: 165–167, Appendix H). Taulukoissa 3, 4 ja 5 näkyvät kaikista kolmesta virkkeestä yhden osallistujaryhmän eli tekstiversiota 2 (ks. taulukko 1) editoineiden ratkaisut.

Taulukko 3: Virke 18, tekstiversiota 2 editoineet

RBMT	Me julkaisemme	teidän nimenne,	kun te toimitatte sitä	(jos te ette kiellä meitä),
P12	Me julkaisemme	teidän nimenne,	kun te toimitatte sitä	(jos te ette kiellä meitä),
P14	Me julkaisemme	nimesi	sellaisena kuin sen toimitat	(jos et kiellä meitä),
P16	Julkaisemme	nimesi	siinä muodossa kuin se on annettu	(ellet kiellä julkaisemasta sitä),
T01	Me julkaisemme	nimesi	kuten sen ilmoitat	(jos et kiellä meitä),
T06	Me julkaisemme	nimesi	kuten se on ilmoitettu	(ellet kiellä meitä),
T08	Julkaisemme	nimesi	kuten olet sen ilmoittanut	(ellet erikseen kiellä),
T12	Me julkaisemme	nimesi	toimittamassasi muodossa	(jos et kiellä meitä),
T13	Me julkaisemme	lähetetyt nimet		(paitsi jos toisin toivotaan),
T16	Me julkaisemme	teidän nimenne,	jos te toimitatte sen	(jos te ette kiellä meitä),

Taulukko 4: Virke 25, tekstiversiota 2 editoineet

NMT	julkaisemme	nimesi	niin et anna sitä	(paitsi jos et kysy meitä),
P12	julkaisemme	nimesi	niin kuin olet ilmoittanut sen	(paitsi jos et halua nimeäsi julkaistavan),
P14	julkaisemme	nimesi	sellaisena kuin olet sen antanut	(paitsi jos pyydät, ettemme tee niin),
P16	julkaisemme	nimesi	siinä muodossa kuin se on annettu	(ellet kiellä meitä julkaisemasta sitä),
T01	julkaisemme	nimesi	kuten sen ilmoitat	(ellet kiellä meitä),
T06	julkaisemme	nimesi	kuten se on ilmoitettu	(ellet kiellä meitä),
T08	julkaisemme	nimesi	kuten olet sen ilmoittanut	(ellet erikseen kiellä),
T12	julkaisemme	nimesi	toimittamassasi muodossa	(paitsi jos et anna meille siihen lupaa),
T13	julkaisemme	nimesi		(paitsi jos toisin toivotaan),
T16	julkaisemme	nimesi		(paitsi jos kiellät),

Taulukko 5: Virke 27, tekstiversiota 2 editoineet

NMT	Julkaisemme	nimenne	niin kuin annatte sen,	paitsi jos pyydätte meitä olemaan.
P12	Julkaisemme	nimesi	niin kuin ilmoitat sen,	paitsi jos et halua nimäsi julkaistavan.
P14	Julkaisemme	nimenne	niin kuin annatte sen,	paitsi jos pyydätte meitä olemaan tekemättä niin.
P16	Julkaisemme	nimesi	siinä muodossa kuin se on annettu	(ellet kiellä meitä julkaisemasta sitä).
T01	Julkaisemme	nimesi	niin kuin ilmoitat sen,	ellet pyydä meitä olemaan julkaisematta sitä.
T06	Julkaisemme	nimesi	niin kuin annat sen,	ellet kiellä meitä.
T08	Julkaisemme	nimesi	kuten olet sen ilmoittanut,	ellet erikseen kiellä.
T12	Julkaisemme	nimenne	niin kuin annatte sen,	paitsi jos pyydätte meitä olemaan tekemättä niin.
T13	Julkaisemme	lähettäjän nimen,		paitsi jos meitä pyydetään olemaan julkaisematta sitä.
T16	Julkaisemme	nimenne,		paitsi jos kiellätte.

Taulukkojen 3–5 editointeja verrattaessa havaitaan, että ainoastaan T08:n editoinnissa kaikki vertaillut segmentit ovat samassa muodossa ja P16:n editoinnit eroavat vain yhden sanan verran virkkeessä 18. Kovinkaan moni ei siis ole hyödyntänyt yhdessä kohdassa editoidun virkkeen kopiointia toiseen kohtaan.

O'Brien (2006: 177) pohtii sitä, mikä vaikutus jälkieditoijan aiemmilla valinnoilla on myöhempään tekstiin. Toistuvissa virkkeen osissa vaikutusta oli selvästi sillä, missä järjestyksessä konekääntimien tuotokset olivat olleet. Editointeja tarkastellessamme havaitsimme, että jotkut editoijat olivat yhtenäistäneet lukijan puhuttelun yksikön 2. persoonaksi, monikon 2. persoonaksi tai persoonattomaksi muodoksi. Kun vertasimme näitä muotoja virkkeissä 18, 25 ja 27, havaitsimme, että niistä 11 editoijasta, joilla oli keskimmaisessä osassa (virkkeissä 10–18) SMT-käännös, jossa lukijaa sinuteltiin, ja tämän jälkeen (virkkeistä 19–27) RBMT-käännös, jossa oli monikko/teitittelymuoto, 8 muutti monikon toisen persoonan yksikön toiseksi. Niistä yhdeksästä editoijasta, joilla virke 18 oli teitittelymuodossa (RBMT-käännin), virke 25 sinuttelumuodossa ja virke 27 teitittelymuodossa (NMT-käännin), neljä yhtenäisti puhuttelun sinutteluksi, yksi muutti sen osin sinutteluksi, osin persoonattomaan tai passiivimuotoon (esimerkiksi ”julkaisemme nimesi (paitsi jos toisin toivotaan)”) ja loput neljä eivät yhtenäistäneet muotoja. Kun virke 18 oli 13 editoijalla teitittelymuodossa (NMT-käännin) ja virkkeet 25 ja 27 sinuttelumuodossa (SMT-käännin), kolme editoijaa yhtenäisti lähes kaikki puhuttelut sinutteluiksi, kaksi teitittelyiksi ja yksi käytti passiivia tai nollapersoonaa (esimerkiksi ”ellei tätä erikseen kiellä”). Sinutteluksi yhtenäistäminen oli siis yleisem-

pää: yhteensä 11 editoijaa muutti kaikki puhuttelumuodot sinutteluiksi, kaksi kaikki puhuttelut teitittelyiksi ja kaksi käytti suurimmassa osassa puhutteluna nollapersoonaa tai passiivia.

6 Lopuksi

Tässä artikkelissa olemme tarkastelleet valintaverkkoanalyysiä menetelmänä, jolla voi vertailla eri jälkieditoijien samasta konekäännöksestä tuottamia versioita. Menetelmä auttaa hahmottamaan, missä kohdissa editoijat päätyvät joko samoihin tai hyvinkin paljon toisistaan eroaviin ratkaisuihin.

Tässä artikkelissa esitimme Campbellin tapaan yksittäisten ilmausten käännös- ja editointivaihtoehtoja ja O'Brienin tapaan kokonaisia virkkeitä. Yhteen esimerkkiin sovelsimme molempia tapoja (ks. kuva 3 ja taulukko 2). Menetelmät eroavat toisistaan siinä, että valintaverkkoanalyysi Campbellin tapaan edellyttää nimensä mukaisesti kääntäjien tai jälkieditoijien tekemien valintojen analysointia ja luokittelua, ja siinä oletetaan, että prosessissa on jonkinlaisia haarakohtia, joissa valintoja tehdään. O'Brienin tapaan laaditut taulukot ovat pikemminkin yhteenveto eri vaihtoehtoista ilman analysointia. Toisaalta voidaan ajatella, että Campbellin tapaan laadittu analyysi sisältää analysoijan (mahdollisesti subjektiivista) tulkintaa, kun taas O'Brienin taulukko esittää vaihtoehdot sellaisina kuin ne ilmenevät.

O'Brienin taulukkomenetelmä vaikuttaakin helpommalta toteuttaa etenkin silloin, kun halutaan vertailla pidempiä segmenttejä, kuten analyysimme virkkeet 18, 25 ja 27 (ks. alaluku 5.3). Toisaalta segmenttejä voi olla hankalaa kohdistaa toisiinsa silloin, kun versioiden sanajärjestys ja syntaksi eroavat toisistaan, kuten virkkeen 11 analyysi osoittaa (ks. taulukko 2). Taulukkomallista on kuitenkin helpompi hahmottaa, kuka teki mitä ja millaisia ”yhdistelmiä” muodostuu, joten se soveltuu tilanteeseen, jossa tavoitteena on eri editoijien tuotosten vertailu. Kun kiinnostuksen kohteena on ensisijaisesti se, miten tietyt kohdat on ratkaistu, Campbellin analyysi on selkeämpi, sillä siitä on esimerkiksi helpompi hahmottaa, kuinka yleinen kukin ratkaisu oli.

Valitsimme tarkasteltavaksi virkkeitä, joissa esiintyi toistoa. Kiinnostava havainto oli, että osallistujat eivät juuri käyttäneet toistuville ilmauksille samoja vastineita, kun kyse oli yksittäisestä sanasta. Sanan *contribution* kohdalla tämä saattoi johtua myös siitä, että mikään konekääntimien tuottamista vastineista ei ollut erityisen onnistunut, joten se ei ehkä jäänyt editoijan mieleen, vaikka sana esiintyikin neljässä eri virkkeessä. Toinen mahdollinen selitys on se, että osallistujia oli kehoitettu hyödyntämään mahdollisimman paljon konekäännöstä eli tekemään siihen mahdollisimman vähän muutoksia, joten he eivät ehkä siksi ryhtyneet vaihtamaan sanaa. Toisaalta ohje ”korjaa lauserakennetta, jos merkitys on väärä tai epäselvä” ei kuitenkaan estänyt osallistujia editoimasta puhuttelumuotoja: havaitsimme, että jopa 15 osallistujaa yhtenäisti puhuttelumuodot toistuvissa lauseissa. Ehkä sinuttelun ja teitittelyn vaihtelu oli joistakin editoijista niin häiritsevää, että se vaikutti rakennevirheeltä.

Analyysistä kävi ilmi, että joissakin kohdissa editoijat olivat yhtenäistäneet tekstiä (esimerkiksi teitittelyn muuttaminen sinutteluksi), joissakin kohdissa taas eivät (*contri-*

bution-sanan erilaiset vastineet tekstin eri kohdissa). Valintaverkkoanalyysi ei anna selitystä siihen, mistä tällainen vaihtelu johtuu, joten siihen tarvittaisiin muita (prosesintutkimuksen) menetelmiä. Tässä artikkelissa ei ollut mahdollista myöskään tutkia sitä, ovatko erilaisia ratkaisuja tuottaneet kohdat sellaisia, joissa kognitiivinen kuormitus näkyisi myös prosessidatassa. Jatkotutkimuksessa olisikin kiinnostavaa tarkastella prosessissa havaittujen taukojen ja uudelleenmuokkausten mahdollista ilmenemistä kyseisissä kohdissa. Tässä tutkimuksessa koehenkilöryhmänä olivat kääntämisen opiskelijat ja jälkieditointitilanne oli rajoitettu sikäli, että esimerkiksi termilähteitä ei ollut mahdollista käyttää. Jatkossa olisikin kiinnostavaa tutkia myös ammattikäytäntäjä sekä prosessia laajempänä kokonaisuutena, jossa mukana on myös muita tietolähteitä.

Kirjallisuuslähteet

Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Aurelie Neveol, Mariana Neves, Martin Popel, Matt Post, Raphael Rubino, Carolina Scarton, Lucia Specia, Marco Turchi, Karin Verspoor & Marcos Zampieri 2016. Findings of the 2016 Conference on machine translation. Teoksessa: *Proceedings of the First Conference on Machine Translation (WMT)*. Stroudsburg, PA: The Association for Computational Linguistics, 131–198.

Campbell, Stuart 2000a. Critical structures in the evaluation of translations from Arabic into English as a Second Language. *The Translator* 6:2, 211–229.

Campbell, Stuart 2000b. Choice Network Analysis in Translation Research. Teoksessa: M. Olohan (eds) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester: St. Jerome, 29–42.

Carl, Michael 2012. Translog – II: A program for recording user activity data for empirical reading and writing research. Teoksessa: Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.) *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. European Language Resources Association, 4108–4112.

Dragsted, Barbara 2012. Indicators of Difficulty in Translation: Correlating Product and Process Data. *Across Languages and Cultures* 13:1, 81–98. <https://doi.org/10.1556/Acr.13.2012.1.5>.

Dranch, Konstantin 2018. Language Services in Finland 2016-2020. Market overview. Esitys Kieli2018-konferenssissa Helsingissä 3.9.2018.

ELIA, EMT, EUATC, FIT, GALA, LIND 2019. *2019 Language Industry Survey – Expectations and Concerns of the European Language Industry*. Saatavissa: <https://www.euatc.org/industry-surveys/item/550-2019-language-industry-survey-report> [viitattu 17.9.2019].

Hale, Sandra Beatriz & Stuart Campbell 2002. The Interaction between Text Difficulty and Translation Accuracy. *Babel* 48:1, 14–33. Saatavissa: <https://doi.org/10.1075/babel.48.1.02hal> [viitattu 17.9.2019].

ISO/DIS 18587 2016. *Translation Services – Post-editing of Machine Translation Output – Requirements*. Geneva: International Organization for Standardization.

Ketola, Anne 2016. An Illustrated Technical Text in Translation Choice Network Analysis as a Tool for Depicting Word-Image Interaction. *Trans-Kom* 9:1, 79–97.

Ketola, Anne 2018. *Word – Image Interaction in Technical Translation: Students Translating an Illustrated Text*. [Väitöskirja.] Tampereen yliopisto. <https://doi.org/10.13140/RG.2.2.18766.15688> [viitattu 17.9.2019].

Koponen, Maarit & Leena Salmi 2017. Post-editing quality: Analysing the correctness and necessity of post-editor corrections. *Linguistica Antverpiensia. New Series* 16:12, 137–148.

Koponen, Maarit, Leena Salmi & Markku Nikulin 2019. A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation* 33: 1–2, 61–90.

Kuhmonen, Kaisamari 2019. Case study: Experimenting with machine translators. Esitys 4th ELRC Conference -konferenssissa 26.11.2019. Saatavissa: http://lr-coordination.eu/4elrc_agenda [viitattu 29.1.2020].

O'Brien, Sharon 2005. Methodologies for measuring the correlations between post-editing effort and machine translatability. *Machine Translation* 19, 37–58.

O'Brien, Sharon 2006. Machine-Translatability and Post-Editing Effort. An Empirical Study Using Translog and Choice Network Analysis. [Väitöskirja.] Dublin City University.

Pavlović, Nataša 2009. More Ways to Explore the Translating Mind: Collaborative Translation Protocols. Teoksessa: Inger M. Mees, Arnt Lykke Jacobsen & Susanne Göpferich (eds) *Behind the Mind. Methods, Models and Results in Translation Process Research*. Fredriksberg: Samfundslitteratur, 81–105.

Schaeffer, Moritz & Michael Carl 2014. Measuring the Cognitive Effort of Literal Translation Processes. Teoksessa: Ulrich Germann, Michael Carl, Philipp Koehn, Germán Sanchis-Trilles, Francisco Casacuberta, Robin Hill & Sharon O'Brien (eds) *Workshop on Humans and Computer-Assisted Translation*. Gothenburg: Association for Computational Linguistics, 29–37. <https://doi.org/10.3115/v1/w14-0306>.

Kirjoittajat

Maarit Koponen toimii yliopistonlehtorina Turun yliopistossa englannin oppiaineessa. Hänen tutkimusaiheensa keskittyvät konekäännöksen ja käännösteknologian käyttöön, erityisesti jälkieditointiin. Lukuvuonna 2019–2020 hän toimii tutkijatohtorina Helsingin yliopistossa MeMAD-hankkeessa, jossa hän tutkii konekäännöksen laatua. Sähköposti: [maarit.koponen \(at\) utu.fi](mailto:maarit.koponen@utu.fi)

Leena Salmi toimii yliopistonlehtorina Turun yliopistossa ranskan oppiaineessa. Hänen tutkimusaiheitaan ovat mm. käännösteknologia, kääntäjien koulutus ja auktorisoitu kääntäminen. Sähköposti: [leena.salmi \(at\) utu.fi](mailto:leena.salmi@utu.fi)

Liite 1: Analysoidut virkkeet kokonaisuudessaan kääntimittäin. Analysoidut ilmaukset on lihavoitu ja samassa virkkeessä peräkkäin esiintyvät ilmaukset esitetty eri väreillä.

	Lähdeteksti	Sääntöpohjainen (RBMT)	Tilastollinen (SMT)	Neuroverkko (NMT)
4	Whether it is breaking news or a featured item, your contribution can make a difference.	Jos se on uutiset tai esitelty kohta, teidän panoksenne voi tehdä eron.	Se, onko se rikkonut uutisia vai välikysymystä, teidän panoksenne voi vaikuttaa.	Onko se Breaking News tai roolihaamo, oma panos voi vaikuttaa.
11	Terms and conditions for sending contributions to the BBC.	Ehdot avustuksien lähettämistä varten BBC:lle.	Ehdoista ja maksuista lähetetään BBC.	Ehtojen ja ehtojen lähettamisestä BBC:lle.
12	Here are the different ways to send in your contributions :	Tässä ovat eri tavat lähettää teidän panoksianne :	Tässä ovat eri tapoja lähettää puheenvuoroista :	Tässä ovat erilaiset tavat lähettää jäsenmaksuissasi :
18	We will publish your name as you provide it (unless you ask us not to) but we will never publish your mobile phone number.	Me julkaisemme teidän nimenne, kun te toimitatte sitä (jos te ette kiellä meitä), mutta me emme koskaan julkaise teidän matkapuhelinnumeroanne.	Julkaisemme nimesi kun se (ellet pyydä meitä), mutta emme koskaan julkaise matkapuhelimesi numero.	Julkaisemme nimenne niin kuin annatte sen (paitsi jos pyydätte meitä olemaan), mutta emme koskaan julkaise kännykkänumeroanne.
21	You can download the BBC News app on iOS or Android devices and send us images or contributions using the buttons in the app.	Te voitte ladata BBC Newsin sovelluksen iOS tai Android-laitteista ja lähettää meille kuvia, tai panokset , käyttäen nappeja sovelluksessa.	Voit ladata BBC News sovellus iOS tai Androidin laitteilla ja läheta meille kuvia tai painikkeita käyttämällä sovelluksella.	Voit ladata BBC News app paikan iOS tai Android laitteille ja lähettää meille kuvia tai osuuksia käyttämällä napit.
25	If we use your material on BBC programmes or online we will publish your name as you provide it (unless you ask us not to) but we will never publish your email address.	Jos me käytämme teidän materiaalianne BBC-ohjelmista, tai verkon kautta me julkaisemme teidän nimenne kun te toimitatte sitä (jos te ette kiellä meitä), mutta me emme koskaan julkaise teidän sähköpostiosoitettanne.	Jos käytämme omia ohjelmia tai BBC materiaalia netissä julkaisemme nimesi kun se (ellet pyydä meitä), mutta emme koskaan julkaise sähköpostiosoitteesi.	Jos käytämme materiaaliasi BBC:n ohjelmissa tai verkossa julkaisemme nimesi niin et anna sitä (paitsi jos et kysy meitä), mutta emme koskaan julkaise sähköpostiosoitteesi.
27	We will publish your name as you provide it, unless you ask us not to .	Me julkaisemme teidän nimenne, kun te toimitatte sitä, jos te ette kiellä meitä.	Julkaisemme nimesi kuin, jos pyydät meitä.	Julkaisemme nimenne niin kuin annatte sen, paitsi jos pyydätte meitä olemaan.